1. A method for searching a collection of items, wherein each item in the collection has a set of properties, comprising the steps of:

obtaining a query composed of a first set of one or more properties; and

obtaining a result based on applying a distance function to one or more of the items in the collection, wherein

the distance function determines a distance between the query and an item in the collection based on the number of items in the collection that are associated with all of the properties in the intersection of the first set of properties and the set of properties for the item.

2. The method of claim 1, further including the step of associating each item in the collection with a set of properties.

3. The method of claim 1, wherein the step of obtaining a result includes identifying result items whose distance from the query is within a first threshold.

4. The method of claim 3, wherein the step of obtaining a result includes ranking the result items according to their distance from the query.

5. The method of claim 3, wherein the threshold is defined as a number of result items.

6. The method of claim 3, wherein the threshold is defined as a distance.

7. The method of claim 1, further including the step of returning the result.

8. The method of claim 1, wherein the step of obtaining a query includes the step of mapping a received query to a set of one or more properties.

9. The method of claim 1, wherein one or more of the properties are binary.

10. The method of claim 1, wherein one or more of the properties are related by a partial order, and wherein, if an item is associated with a property, then the item is also associated with all ancestors of that property in the partial order.

11. The method of claim 6, wherein one or more of the properties represent numerical values or ranges, and wherein the partial order reflects a set of containment relationships among the numerical values or ranges.

12. The method of claim 1, wherein the properties are grouped into equivalence classes.

13. The method of claim 12, further including the step of grouping the properties into equivalence classes using clustering.

14. The method of claim 13, wherein each property has a set of subproperties, wherein the clustering is performed such that the distance between two properties in the collection is correlated to the number of properties in the collection that are associated with all of the subproperties common to both properties.

15. The method of claim 1, wherein the query corresponds to a single item in the collection.

16. The method of claim 1, wherein the query corresponds to a plurality of items in the collection.

17. The method of claim 1, wherein the query is independent of the items in the collection.

18. The method of claim 1, wherein the step of obtaining a result is constrained to a subcollection of the items in the collection.

19. The method of claim 18, wherein the subcollection is specified as an expression of properties.

20. The method of claim 19, wherein the expression includes a subset of the set of properties that compose the query.

21. The method of claim 1, wherein the step of obtaining a query includes identifying certain properties to be ignored in the step of obtaining a result.

22. The method of claim 1, wherein the distance function is applied explicitly.

23. The method of claim 1, wherein the distance function is applied implicitly.

24. The method of claim 23, wherein the step of obtaining a result includes the step of iterating a random walk process to select potential result items.

25. The method of claim 24, wherein the step of obtaining a result includes ranking the potential result items by frequency and selecting the potential result items having higher frequencies.

26. The method of claim 23, wherein the step of obtaining a result includes iterating through one or more subsets of the query and identifying items associated with the one or more subsets.

27. The method of claim 26, wherein the one or more subsets are prioritized according to the number of items in the collection that have all of the properties in each subset and wherein iterating through one or more subsets of the query is continued until a first threshold is reached.

28. The method of claim 1, wherein the step of obtaining a result includes applying a Euclidean distance function.

29. The method of claim 28, wherein the step of obtaining a result includes merging a first result determined by applying the distance function and a second result determined by applying the Euclidean distance function.

30. The method of claim 28, wherein the step of obtaining a result includes determining a first result by applying either the distance function or the Euclidean distance function and applying the other distance function to the first result.

31. A method for analyzing two sets of properties from a plurality of sets of properties, comprising the steps of:

determining a set of common properties in the intersection of the two sets of properties;

determining the number of sets of properties from the plurality of sets of properties that include the set of common properties; and

assessing the distance between the two sets of properties as a function of the number of sets of properties that include the set of common properties.

32. A method for analyzing the relationship between two items in a collection of items, wherein each item in the collection is associated with a set of properties, comprising the steps of:

obtaining a set of properties with which the two items are commonly associated; and

determining the degree of commonality between the two items as a function of the number of items in the collection that are associated with all of the properties with which the two items are commonly associated.

33. A computer program product, residing on a computer readable medium, for use in searching a collection of items, the computer program product comprising instructions for causing a computer to:

receive a query composed of one or more properties; and

obtain a result based on applying a distance function to one or more items in the collection, wherein

the distance function determines a distance between the query and an item in the collection based on the number of items in the collection that are associated with all of the properties in the intersection of the first set of properties and the set of properties for the item.

34. The computer program product of claim 33, wherein the instructions cause the computer to obtain a result by identifying exactly the items whose distance from the query is within a threshold.

35. The computer program product of claim 33, wherein the instructions cause the computer to obtain a result by identifying approximately the items whose distance from the query is within a threshold according to a heuristic.

36. The computer program product of claim 35, wherein the heuristic permits a trade-off between the accuracy and the performance of a search.

37. The computer program product of claim 35, wherein the heuristic includes the use of a random walk process.

38. A computer system for managing data records comprising:

an information retrieval subsystem that stores and retrieves data records, each data record being associated with a set of properties; and

a similarity search subsystem that receives similarity search queries and processes similarity search queries based on a distance function, a similarity search query being associated with a first set of properties, wherein

the distance function determines a distance between the query and a data record in the collection based on the number of data records in the collection that are associated with all of the properties in the intersection of the first set of properties and the set of properties for the data record.

39. The computer system of claim 38, further including a clustering subsystem that employs the distance function of the similarity search subsystem to construct a graph.

40. A method for applying a matching algorithm to a collection of items, each item being associated with a set of properties, comprising the steps of:

constructing a graph having nodes that correspond to items, and having edges that correspond to pairs of items, wherein each edge has a cost correlated to the number of items in the collection that are associated with all of the properties in the intersection of the sets of properties for the two items that the edge links; and

identifying a subset of the edges that constitutes a minimum-cost matching with respect to the graph.

41. A method for applying a clustering algorithm to a collection of items, each item being associated with a set of properties, comprising the steps of:

constructing a graph having nodes that correspond to items, and having edges that correspond to pairs of items, wherein each edge has a cost correlated to the number of items in the collection that are associated with all of the properties in the intersection of the sets of properties for the two items that the edge links; and

identifying a collection of subsets of the edges that constitutes a minimum-cost clustering with respect to the graph.